

501.42942X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): KOIZUMI, et al.

Serial No.: Not assigned

Filed: July 22, 2003

Title: INFORMATION PROCESSOR AND PROGRAM FOR
IMPLEMENTING INFORMATION PROCESSOR

Group: Not assigned

LETTER CLAIMING RIGHT OF PRIORITY

Mail Stop Patent Application
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

July 22, 2003

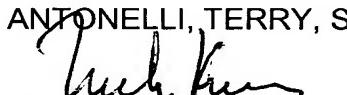
Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby
claim(s) the right of priority based on Japanese Application No.(s) 2002-341671 filed
November 26, 2002.

A certified copy of said Japanese Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP


Melvin Kraus
Registration No. 22,466

MK/amr
Attachment
(703) 312-6600

310201540
HK

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office

出願年月日

Date of Application:

2002年11月26日

出願番号

Application Number:

特願2002-341671

[ST.10/C]:

[JP2002-341671]

出願人

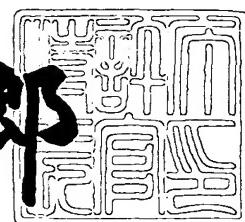
Applicant(s):

株式会社日立製作所

2003年 4月25日

特許庁長官
Commissioner,
Japan Patent Office

太田 信一郎



出証番号 出証特2003-3030724

【書類名】 特許願
【整理番号】 H02015401A
【あて先】 特許庁長官 殿
【国際特許分類】 G06F 17/30
【発明者】
【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所 中央研究所内
【氏名】 小泉 敦子
【発明者】
【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所 中央研究所内
【氏名】 森本 康嗣
【発明者】
【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所 中央研究所内
【氏名】 順井 裕之
【発明者】
【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所 中央研究所内
【氏名】 秋良 直人
【特許出願人】
【識別番号】 000005108
【氏名又は名称】 株式会社日立製作所
【代理人】
【識別番号】 100075096
【弁理士】
【氏名又は名称】 作田 康夫
【電話番号】 03-3212-1111

【手数料の表示】

【予納台帳番号】 013088

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 情報処理装置、該情報処理装置を実現するためのプログラム

【特許請求の範囲】

【請求項1】

複数のデータを格納する記憶手段と、

上記記憶されるデータのうち単語若しくは語を共通に有するデータに共通の属性を付する手段と、

上記データを分析する解析手段とを有し、

上記解析手段は、属性付けのなされていないデータにはネガティブ語辞書を用いた分析を行い、上記属性付けのなされているデータには異なる分析を行うことを特徴とする情報処理装置。

【請求項2】

上記情報処理装置は、

入力手段と、

上記入力手段を介して受けつけたキーワードを用いて上記データベース内を検索する手段を有し、

上記属性を付する手段は、上記検索の結果抽出されたデータにその旨の属性付けを行うことを特徴とする請求項1記載の情報処理装置。

【請求項3】

上記入力手段は上記検索手段において抽出された回数の指定を受け、

上記解析手段は、上記回数以下抽出された旨の属性を有するデータと、上記回数より多い回数抽出された旨の属性を有するデータとで異なる解析方法で分析を行うことを特徴とする請求項2記載の情報処理装置。

【請求項4】

上記ネガティブ語辞書は、漢字単位の語を格納する第1の辞書と該漢字を含む單語を格納する第2の辞書とから構成され、

上記解析手段は、上記データから上記第1及び第2の辞書に格納される語を検索し、上記第1の辞書に格納される漢字を含むとして検索された单語のうち上記第2の辞書にないものを上記表示手段に表示して、該表示した单語のうち指定され

た単語を上記第2の辞書に格納することを特徴とする情報処理装置。

【請求項5】

モダリティを表現する単語を格納する辞書をさらに有し、

上記解析手段は、上記辞書を使った分析を行うことを特徴とする請求項1乃至4の何れかに記載の情報処理装置。

【請求項6】

上記記憶されるデータから単語と単語の関連度を計算する手段と、

上記記憶されるデータから重要タームを抽出する手段と、

上記関連度の情報を用いて上記重要タームをクラスタリングしシソーラス概観を生成する手段と、

上記生成されたシソーラス概観を表示手段に表示する手段とを有し

上記表示手段は、上記入力手段を介して選択された上記シソーラス概観のクラスタに属する重要タームを表示し、

上記表示される重要タームのうち上記指示入力手段を介して指示された重要タームを上記キーワードとして設定することを特徴とする請求項2乃至5の何れかに記載の情報処理装置。

【請求項7】

漢字単位の語を格納する第1の辞書と、

該漢字を含む単語を格納する第2の辞書と、

表示手段と、

入力手段と、

記録手段に記録されるデータから上記第2の辞書に格納される単語を検索する手段を有し、

上記検索手段は、上記第1の辞書に格納される漢字を含む単語も検索し、上記第1の辞書に格納される漢字を含むとして検索された単語を上記表示手段に表示して、該表示した単語のうち指定された単語を上記第2の辞書に格納することを特徴とする情報処理装置。

【請求項8】

上記指定されなかった単語を蓄積する第3の辞書を有することを特徴とする請求

項7記載の情報処理装置。

【請求項9】

上記第1の辞書は否定的な意味を持つ漢字を格納し、

上記第2の辞書は、否定的な意味をもつ単語を格納することを特徴とする請求項7又は8に記載の情報処理装置。

【請求項10】

キーワードの入力を受けるステップと、

複数のデータを格納する記憶手段に格納される複数のデータを上記キーワードを用いて検索するステップと、

上記検索の結果抽出されたデータに共通の属性を付するステップと、

上記属性付けのされていないデータについてネガティブ語辞書を用いた分析を行い、上記属性付けのなされているデータには上記ネガティブ辞書とは異なるデータを用いた分析を行うステップとをコンピュータに実行させることを特徴とするプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、自然言語で記述されたテキストから知識を抽出するテキスト分析方法に関する。主として、コールセンターの応答履歴の分析を対象とする。

【0002】

【従来の技術】

ユーザが指定したキーワードにより文書を分類する文書分類システムとしては、文書中の単語の出現頻度に基づいて未使用視点（まだ分類に使っていないキーワード）を検出し表示することによりキーワードによる分類を支援する文書分類システムがある（例えば、特許文献1参照）。

リスク管理の上で有用な知識を抽出する手段としては、「失礼」「失望」などのネガティブな表現に着目することが考えられる。ネガティブ表現を抽出する方法としては、ドメインに応じて「失注」、「苦情」などのネガティブな意味を持つキーワードを予めセットしておき、検索を実行して、ヒットした場合にはアラ

トを出すという方法が考えられる。更に、文書分類のためのキーワード辞書をユーザが更新する手段を設けた文書分類システムもある（例えば、特許文献2参照）。

【特許文献1】特開2001-101226号公報

【特許文献2】特開2001-184351号公報

【発明が解決しようとする課題】

従来のキーワードによる文書分類技術は、高頻度知識の抽出・分類に適しているが、コールセンターの応答履歴からリスク管理上有用な情報や顧客の生の声を抽出するには、低頻度の知識の抽出が重要課題である。すなわち、大量のありふれた情報を取り除いた中から、効率よく、かつ漏れなく、真に有用な知識を抽出する必要がある。本発明の目的は、高頻度の問合せに基づいてFAQを作成することと、低頻度の問合せの中からリスク管理上有用な情報を抽出することにある。

リスク管理の目的でテキスト分析を行う際に、ネガティブな表現を抽出することが考えられる。ネガティブな表現を抽出するためには、ドメインに応じて「失望」「失礼」などのキーワードをセットしておき、検索を実行する方法が考えられるが、予めキーワードを設定することに手数がかかる上に、網羅することが困難であり、漏れが多く発生するという問題がある。

【0003】

【課題を解決するための手段】

上記課題を解決するため、テキスト分析支援システムにおいて、低頻度情報を抽出するための手段として、高頻度情報を含む文書を抽出してフォルダに保存した後、残りの文書を集めて低頻度情報のフォルダに保存する機能を設け、低頻度情報のフォルダのデータにはネガティブ表現の抽出漏れとノイズをなくすための手段として、「失」「負」などのネガティブな意味を持つ文字を格納した辞書を用いて対象テキストからネガティブ語候補を抽出し、ネガティブ語と判定したものをネガティブ語辞書に登録した上で、ネガティブ語辞書を用いてネガティブ表現の抽出を行うようとする。

また、

【0004】

【発明の実施の形態】

以下、本発明の実施例について説明する。本実施例は、コールセンタの応答履歴を対象としたテキスト分析支援システムである。以下、図面を使って詳細に説明する。

(システム構成)

図1は本発明の第1の実施例を示すテキスト分析支援システムの構成図である。本システムは、CPU101、入力装置102、表示装置103、コールセンタ応答履歴データベース104、シソーラスブラウジング用データ格納部105、文書保存フォルダ106、低頻度知識抽出用データ格納部107、メモリ108によって構成されている。シソーラスブラウジング用データ格納部105は、関連シソーラス格納部1051、タームベクトル格納部1052、およびシソーラス概観格納部1053によって構成されている。低頻度知識抽出用データ格納部107は、ネガティブ表現抽出機能を実現するためのネガティブ文字辞書1071、ネガティブ語辞書1072、ネガティブ語トップワード辞書1073、モダリティ表現抽出機能を実現するためのモダリティ表現辞書1074、モダリティ表現トップワード辞書1075によって構成されている。メモリ108には、シソーラスブラウジング用データ生成処理手段1081、シソーラスブラウジング処理手段1082、文書検索手段1083、ネガティブ語候補抽出手段1084、ネガティブ語辞書作成手段1085、モダリティ表現候補抽出手段1086、モダリティ表現辞書作成手段1087が記憶されている。

(コールセンタ応答履歴データベース)

図2にコールセンタ応答履歴データベース104のデータ構造を示す。コールセンタ応答履歴データベース104の各レコードには、問合せID1041、応答履歴メモ1042、キーワード検索で検索済みであることを示す検索フラグ1043、分類フォルダに分類済みであることを示す分類フラグ1044が記述されている。

(シソーラスブラウジング機能)

本システムは、高頻度情報を含む文書の抽出を支援するシソーラスブラウジ

グ機能を備えている。ここでいうシソーラスとは、文書群中の特徴的な単語とその関係を示すネットワーク表現である。本システムのシソーラスブラウジング機能は、文書群からシソーラスを自動生成する機能と、生成したシソーラスの概観や細部を表示する機能（概観表示・ズーム表示）からなる。シソーラス自動生成およびシソーラス表示は、例えば特開2000-227917に記載されているシソーラスブラウジング方法によって行う。以下、本システムにおいてシソーラスブラウジング機能を実現するためのデータおよび処理手順の概要を説明する。まず、シソーラスブラウジング機能を実現するためのデータについて説明する。シソーラスブラウジング用データ格納部105は、関連シソーラス格納部1051、タームベクトル格納部1052、およびシソーラス概観格納部1053によって構成されている。

関連シソーラス格納部1051には、コールセンタ応答履歴データベース104の応答履歴メモ1042に格納された文書データから生成した関連シソーラスが格納されている。関連シソーラスとは、単語と単語の関連度を示すものである。本実施例では、関連度は2つの単語の共起しやすさを表すものであり、それぞれの単語の頻度と共起頻度（文書中のある範囲内に2つの語が同時に出現する頻度）に基づいて計算される。図3に関連シソーラス格納部1051のデータ構造を示す。関連シソーラス格納部1051は、レコードID10511、タームX10512、タームY10513、および関連度10514から構成される。タームX10512およびタームY10513には、関連関係にあるタームを、関連度10514にはその関連度を格納する。

タームベクトル格納部1052には、コールセンタ応答履歴データベース104の応答履歴メモ1042に格納された文書データから抽出したタームベクトルが格納されている。タームベクトルとは、文書を特徴付けるタームのリストであり、「Salton, G., et al.: A Vector Space Model for Automatic Indexing, Communications of the ACM, Vol.18, No.11(1975).」に記載のtf-idf法（Term Frequency inverse Document Frequency）を利用することにより抽出可能である。このtf-idf法は、文書インデクシング方法として最もよく知られているもののひとつであり、ある文書におけるタームの出現頻度（tf）と、当該タームが出現

した文書数の逆数(idf)をかけた値を当該文書におけるタームの重要度とし、当該文書において重要度の高いターム（すなわち重要ターム）を抽出してタームベクトルとする技術である。図4にタームベクトル格納部1052のデータ構造を示す。タームベクトル格納部1052は、レコードID10521、問合せID10522および重要タームリスト10523から構成される。問合せID10521には、コールセンタ応答履歴データベース104に格納された応答履歴のIDを格納し、重要タームリスト10522には当該応答履歴の応答メモに出現するタームのうち重要なもののリストが格納される。

シソーラス概観格納部1053には、関連シソーラス格納部1051に格納された関連シソーラスの概観が格納されている。シソーラス概観とは、文書群中のもっとも特徴的な単語を代表タームとして抽出し、関係の強い代表タームをタームクラスタとしてまとめたものである。図5にシソーラス概観格納部1053のデータ構造を示す。シソーラス概観格納部1053は、タームグループ番号10531およびタームリスト10532から構成される。タームリスト10532には、タームクラスタに属するタームのリストが格納される。

【0005】

以上、シソーラスブラウジング用データについて説明した。

次に、シソーラスブラウジング機能を実現するためのシソーラスブラウジング用データ生成処理手順および、シソーラスブラウジング処理手順について図7および図8のフローチャートを用いて説明する。

（シソーラスブラウジング用データ生成処理手順）

まず、分析環境準備として、シソーラスブラウジング用データを作成する。図7に示すように、シソーラスブラウジング用データ生成処理では、まず文書データからタームとタームの関連度を示す関連シソーラスを生成し（ステップ701）、各文書のタームベクトルを抽出して（ステップ702）、シソーラス概観を生成する（ステップ703）。シソーラス概観は、文書群中のもっとも特徴的な単語を代表タームとして抽出し、関係の強い代表タームをタームクラスタとしてまとめたものである。代表ターム抽出処理では、各文書タームベクトルを構成する

重要タームのうち、多くの文書で重要タームとなったタームを代表タームとする。タームクラスタ生成処理では、関連シソーラスに格納されたターム間の関連度に基づいて関連度の高い代表タームをひとつのクラスタにまとめる。

(シソーラスブラウジング処理手順)

図8に示すように、シソーラスブラウジング処理では、まずシソーラス概観格納部1053に格納されたシソーラス概観を例えば図6のシソーラス概観表示部602に示すような形でユーザに表示する(ステップ801)。シソーラス概観表示部602は、タームリスト表示部6021および選択ボタン6022からなる。タームリスト表示部6021には、シソーラス概観格納部1053に格納されているタームリスト10532が表示される。次にユーザがタームクラスタリスト6021を選択ボタン等の指示入力手段6022で選択してズームボタン6033でズームを指示すれば(ステップ802)、ユーザが選択したタームクラスタに属するタームの関連タームを関連シソーラス1051より取得する(ステップ803)。そして、それらをクラスタリングし(ステップ804)、生成したタームクラスタを関連タームクラスタ表示部604に表示する(ステップ805)。ユーザからのシソーラスブラウジング終了の指示があれば(ステップ806)、処理を終了し、なければステップ802の処理に戻る。ステップ802のズーミング指示において、関連タームクラスタ表示部604に表示されているタームクラスタ6041を選択ボタン6042で選択してズームボタン6033でズームを指示すれば、該関連タームクラスタの関連語が関連タームクラスタ表示部604に表示される。また、シソーラス概観表示部602あるいはタームクラスタ表示部604に表示されているタームをクリックしてからズームボタン6033をクリックすると、該タームの関連語が関連タームクラスタ表示部604に表示される。ユーザは、関連クラスタ数6031およびクラスタ内ターム数6033を選択することにより、いくつのクラスタに分けるか、1つのクラスタについて何ターム抽出するかを指定することができる。

(シソーラスブラウジングによる効果)

このようにキーワードで文書を検索する機能と、検索した文書をフォルダに保存する機能を設け、ユーザがキーワードとして入力した語に関連する問合せを抽出し、FAQ作成のために保存することができるようになる。また、応答履歴全体からシソーラスを生成し、シソーラスの全体構造を示すシソーラス概観から、ユーザが選択したタームを含む部分構造へと、ユーザをナビゲートするシソーラスブラウジング機能を設け、ユーザがキーワードを想起しやすいようにする。シソーラス概観を眺めることにより、文書群中のトピックを俯瞰することができる。1つのタームクラスタにまとめられた代表タームの並びを見ると、トピックやその内容を推測することができる。タームの関連語をクラスタ表示（関係の強い語をタームクラスタとしてまとめて表示）することにより、タームに対応するトピックのサブトピックとその内容を推測することができる。

【0006】

本システムは、シソーラスブラウジング機能およびキーワード文書検索機能により高頻度情報を含む文書を抽出して分類フォルダに保存した後、残りの文書を集めて低頻度情報のフォルダに保存する機能を備えている。図6に文書分類操作画面の構成を示す。図6に示すように、文書分類操作画面601は、シソーラスブラウジング機能のためのシソーラス概観表示部602、シソーラスズーミング指示部603、関連タームクラスタ表示部604、キーワード文書検索機能のための文書検索指示部605、文書検索結果表示部606、文書分類保存機能のための文書保存部607からなる。

シソーラス概観表示部602は、タームリスト表示部6021および選択ボタン6022からなる。タームリスト表示部6021には、シソーラス概観格納部1053に格納されているタームリスト10532が表示される。シソーラスズーミング指示部603は、クラスタ数6031、クラスタ内ターム数6032、ズームボタン6033からなる。

関連タームクラスタ表示部604は、タームリスト表示部6041および選択ボタン6042からなる。

文書検索指示部605は、検索ターム入力部6051および検索ボタン6052からなる。文書検索結果表示部606は、文書表示部6061および文書選択ボ

タン6062からなる。文書保存部607はフォルダ名表示部6071およびフォルダ選択ボタン6072からなる。

(文書分類手順)

本システムは、高頻度情報を含む文書を抽出してフォルダに保存した後、残りの文書を集めて低頻度情報のフォルダに保存する機能を備えている。図9は、本システムによる文書分類手順を示すフローチャートである。本システムによる文書分類手順について、図6の文書分類操作画面および図9のフローチャートを用いて説明する。まず、分類開始指示があると（ステップ901）、コールセンタ応答履歴データベース104にアクセスし、検索済みであることを示す検索フラグ1043と、分類済みであることを示す分類フラグ1044の値を“0”にリセットする。ユーザがターム入力部6051にタームを入力し、検索ボタン6052をクリックしてキーワード文書検索を指示すると（ステップ903）、コールセンタ応答履歴データベース104の応答履歴メモ1042を対象にキーワード文書検索を行い（ステップ904）、コールセンタ応答履歴データベース104の検索フラグ1043に検索済みであることを示すフラグ“1”を設定し（ステップ905）、文書検索結果を文書検索結果表示部606の文書表示部6061に表示する（ステップ906）。ユーザが文書検索結果一覧から保存したい文書を選択して文書選択ボタン6062とフォルダ選択ボタン6072をクリックすると（ステップ907）、選択された文書を文書保存フォルダ106へ保存し（ステップ908）、コールセンタ応答履歴データベース104の分類フラグ1044に分類済みであることを示すフラグ“1”を設定する（ステップ909）。ユーザから分類終了の指示があれば（ステップ910）、検索済みフラグ=0の文書を低頻度文書フォルダに保存する（911）。

低頻度文書フォルダへの文書保存方法の代案としては、分類済みフラグ=0の文書を低頻度文書フォルダに保存するようにしてもよい。また、文書保存フォルダに選択フラグを用意し、ユーザが指定したフォルダに分類済みの文書以外の文書を低頻度文書フォルダに保存するようにしてもよい。さらに、検索済み、分類済みかどうかを示す検索フラグおよび分類済みフラグの変わりに検索回数および分類回数を更新するようにし、検索回数あるいは分類回数が閾値よりも低いものを

低頻度文書フォルダに保存するようにしてもよい。

【0007】

本システムは、キーワード想起を支援するシソーラスブラウジング機能を備えている。ユーザは、シソーラスブラウジングの過程で、表示されたタームを選択することによりキーワード文書検索を行うこともできる。シソーラス概観表示部602のタームリスト表示部6021に表示されたタームをクリックすると該タームが検索ターム入力部6051にコピーされる。また、シソーラス概観表示部602の選択ボタン6022をクリックすると、タームリスト表示部6021に表示されている全てのタームが検索ターム入力部6051にコピーされる。同様に、関連タームクラスタ表示部604のタームリスト表示部6041に表示されたタームをクリックすると該タームが検索ターム入力部6051にコピーされ、選択ボタン6042をクリックすると、タームリスト表示部6041に表示されている全てのタームが検索ターム入力部6051にコピーされる。シソーラスには、応答履歴全体に出現するタームが関連付けて格納されている。したがって、シソーラスブラウジングをすることにより、高頻度情報を収集・分類することができる。

(低頻度情報からの知識抽出)

以上に述べたように、本システムでは、分類開始から終了までの間に一度も検索されていな文書、あるいは、どの分類フォルダにも分類されていない文書をまとめて低頻度情報フォルダに格納することができる。リスク管理の目的でテキスト分析を行う際に、「失礼」「失望」などのネガティブな意味を持つ単語や、「くれないのか」「そもそも」「なんなのか」「欲しい」などのモダリティ表現が有効な手がかりとなる。そこで、低頻度情報からリスク管理上有用な知識を抽出する手段として、ネガティブな表現を抽出する機能と、顧客やオペレータの心的態度を表すモダリティ表現を抽出する機能を設ける。以下、低頻度情報フォルダに保存された応答履歴メモからネガティブ表現およびモダリティ表現を含む文書を抽出する手順の概要を図21のフローチャートに従って説明する。まず、低頻度情報フォルダに保存された応答履歴メモから、ネガティブ語候補・モダリティ表現候補を抽出する（ステップ2101）。次に、ネガティブ語候補・モダリティ表

現候補のうち、ユーザが選択したものをネガティブ語辞書・モダリティ表現辞書に登録する（ステップ2102）。最後に、低頻度情報フォルダの文書に対して、ネガティブ語辞書およびモダリティ表現辞書に登録された語をキーワードとしてキーワード検索を行うことにより（ステップ2103）、ネガティブ語およびモダリティ表現を含む文書を抽出し、内容を確認する（ステップ2104）。

以下、ネガティブ表現およびモダリティ表現の抽出の手順について詳細に述べる。

（ネガティブ表現の抽出）

応答履歴メモからネガティブな表現を抽出する手段として、本システムは、応答履歴メモからネガティブ語候補を抽出するネガティブ語候補抽出機能と、ネガティブ語候補の中でユーザがネガティブ語と判定した語をネガティブ語辞書に登録するネガティブ語辞書作成機能とを備えている。これらの機能を実現するため、本システムは、「失」「負」「遅」などのネガティブ語の構成要素となりやすい文字を登録したネガティブ文字辞書1071、ネガティブ語であることが判定済みの語が登録されているネガティブ語辞書1072、ネガティブ語でないことが判定済みの語が登録されているネガティブ語トップワード辞書1073を備えている。

図12に、ネガティブ文字辞書1071のデータ構造を示す。ネガティブ文字辞書の各レコードには、レコードID10711、ネガティブ文字10712、ネガティブ度10713、ネガティブ語辞書登録語数10714、ネガティブ語ストップワード辞書登録語数10715が記述されている。ネガティブ語辞書登録語数10714は、ネガティブ語辞書に登録されている単語のうち、当該ネガティブ文字を含む単語の語数である。ネガティブ語ストップワード辞書登録語数10715は、ネガティブ語ストップワード辞書1073に登録されている単語のうち、当該ネガティブ文字を含む単語の語数である。ネガティブ度10713には、ネガティブ語候補として抽出された単語のうちネガティブ語辞書に登録された単語の割合を示す0～1の値が記述されている。あるいは、ネガティブ度の値はユーザが任意に設定するようにしてもよい。図13に、ネガティブ語辞書10

72のデータ構造を示す。ネガティブ語辞書の各レコードには、レコードID10721、ネガティブ語10722、ネガティブ度10723が記述されている。ネガティブ度10723には、ネガティブ文字辞書に記述されたネガティブ度10713の値が記述されている。図14に、ネガティブ語トップワード辞書1073のデータ構造を示す。ネガティブ語トップワード辞書の各レコードには、レコードID10731、ネガティブ語トップワード10732が記述されている。

以下、ネガティブ語候補抽出の手順を図17のフローチャートにしたがって説明する。まず、応答履歴メモ1042にあらわれるすべての単語を抽出し、単語リストを作成する（ステップ1701）。単語リストの単語を1語読み（ステップ1703）、ネガティブ文字辞書1071を参照し、ネガティブ文字を含むかどうかを判定する（ステップ1704）。ネガティブ文字を含む場合は、ネガティブ語辞書1072を参照し、ネガティブ語辞書1072に登録済みであるかどうかを判定する（ステップ1705）。ネガティブ語辞書1072に登録済みの場合は、ネガティブ語であることがすでにわかっているので、ネガティブ語候補として抽出せずにこの単語に関する処理を終了する。ネガティブ語辞書1072に未登録の場合は、ネガティブ語トップワード辞書1703を参照し、ネガティブ語トップワード辞書1073に登録済みであるかどうかを判定する（ステップ1706）。ネガティブ語トップワード辞書1073に登録済みの場合は、ネガティブ語でないことがすでにわかっているので、ネガティブ語候補として抽出せずにこの単語に関する処理を終了する。そして、ネガティブ語辞書にもネガティブ語トップワード辞書にも登録されていない単語をネガティブ語候補リストに登録する（ステップ1707）。単語リストに登録されているすべての単語について同様の処理を行うことにより、ネガティブ文字を含む単語のうち、ネガティブ語辞書にもネガティブ語トップワード辞書にも登録されていない単語をネガティブ語候補リストに登録する。

以下、ネガティブ語辞書作成の手順を図18のフローチャートにしたがって説明する。まず、ネガティブ語候補に対してネガティブ語かどうかの判定を行うため、ネガティブ語候補リストを画面に表示する（ステップ1801）。図11にネ

ガティブ語判定画面の表示例を示す。ネガティブ語判定画面には、ネガティブ語候補表示部11011、ネガティブ語辞書既登録語表示部11012、ネガティブ語トップワード辞書既登録語表示部11013、登録ボタン11014が配置されている。ネガティブ語辞書既登録語表示部11012およびネガティブ語トップワード辞書既登録語表示部11013は判定のための参考情報として表示するものだが、省いても良い。ユーザは、ネガティブ語候補表示部11011に表示されたネガティブ語候補に対してネガティブ語かどうかを判定し、ネガティブ語と判定した語にチェックマークをいれる（ステップ1802）。ユーザが登録ボタン11014をクリックすると（ステップ1803）、ネガティブ語と判断された語がネガティブ語辞書に登録される（ステップ1804）。ネガティブ語と判断されなかった語は、ネガティブ語トップワード辞書に登録される（ステップ1805）。

（モダリティ表現の抽出）

次に、顧客やオペレータの心的態度を表すモダリティ表現を抽出する機能について述べる。図15に、モダリティ表現辞書1074のデータ構造を示す。モダリティ表現辞書の各レコードには、レコードID10741、モダリティ表現10742、品詞10743、モダリティ10744が記述されている。図16に、モダリティ表現トップワード辞書1075のデータ構造を示す。モダリティ表現トップワード辞書の各レコードには、レコードID10751、モダリティ表現トップワード10752、品詞10753が記述されている。

【0008】

以下、モダリティ表現候補抽出の手順を図19のフローチャートにしたがって説明する。まず、応答履歴メモ1042にあらわれるすべての単語を抽出し、単語リストを作成する（ステップ1901）。単語リストの単語を1語読み（ステップ1903）、品詞が副詞か助動詞の場合は（ステップ1904）、モダリティ表現候補抽出の処理を進める。すなわち、モダリティ表現辞書1074を参照し、モダリティ表現辞書1074に登録済みであるかどうかを判定する（ステップ1905）。モダリティ表現辞書1074に登録済みの場合は、モダリティ表現であることがすでにわかっているので、モダリティ表現候補として抽出せずに

この単語に関する処理を終了する。モダリティ表現辞書1074に未登録の場合は、モダリティ表現ストップワード辞書1705を参照し、モダリティ表現ストップワード辞書1075に登録済みであるかどうかを判定する（ステップ1906）。モダリティ表現ストップワード辞書1075に登録済みの場合は、モダリティ表現でないことがすでにわかっているので、モダリティ表現候補として抽出せずにこの単語に関する処理を終了する。そして、モダリティ表現辞書にもモダリティ表現ストップワード辞書にも登録されていない単語をモダリティ表現候補リストに登録する（ステップ1907）。単語リストに登録されているすべての単語について同様の処理を行うことにより、品詞が副詞あるいは助動詞である単語のうち、モダリティ表現辞書にもモダリティ表現ストップワード辞書にも登録されていない単語をモダリティ表現候補リストに登録する。

以下、モダリティ表現辞書作成の手順を図20のフローチャートにしたがって説明する。まず、モダリティ表現候補に対してモダリティ表現かどうかの判定を行うため、モダリティ表現候補リストを画面に表示する（ステップ2001）。モダリティ表現判定画面は、図11のネガティブ語判定画面と同様のものを用いる。ユーザは、画面に表示されたモダリティ表現候補に対してモダリティ表現かどうかを判定し、モダリティ表現と判定した語にチェックマークをいれる（ステップ2002）。ユーザが登録ボタンをクリックすると（ステップ2003）、モダリティ表現と判断された語がモダリティ表現辞書に登録される（ステップ2004）。モダリティ表現と判断されなかった語は、モダリティ表現ストップワード辞書に登録される（ステップ1805）。

【0009】

【発明の効果】

本発明によれば、応答履歴メモに含まれる情報を高頻度情報と低頻度情報に分けることができ、それぞれに適したテキスト分析方法を適用することができるという効果がある。高頻度情報に対しては、トピックで分類することにより、FAQ作成支援に活用することができる。低頻度情報に対しては、ネガティブ表現およびモダリティ表現というトピックとは別の観点から、リスク管理上有用な知識を抽出することができる。

本発明のネガティブ表現抽出方法によれば、文字を手がかりにして分析対象テキストに含まれるネガティブ語候補を抽出するので、抽出漏れを防ぐことができる。抽出したネガティブ語候補についてネガティブ語かどうかの判定を人手で行う必要があるが、ネガティブ語かどうか判定済みの語をネガティブ語辞書およびネガティブ語トップワード辞書に蓄積していくので、繰り返すうちにネガティブ語候補として抽出されるものが減っていくという効果がある。

【図面の簡単な説明】

【図1】

本発明のテキスト分析支援システムの実施例のシステム構成図である。

【図2】

コールセンター応答履歴データベースのデータ構造を示す図である。

【図3】

関連シソーラス格納部のデータ構造を示す図である。

【図4】

タームベクトル格納部のデータ構造を示す図である。

【図5】

シソーラス概観格納部のデータ構造を示す図である。

【図6】

文書分類操作画面の構成を示す図である。

【図7】

シソーラスブラウジング用データ生成処理手順を示すフローチャートである。

【図8】

シソーラスブラウジング処理手順を示すフローチャートである。

【図9】

文書分類手順を示すフローチャートである。

【図10】

文書保存フォルダのデータ構造を示す図である。

【図11】

ネガティブ語判定画面の表示例を示す図である。

【図12】

ネガティブ文字辞書のデータ構造を示す図である。

【図13】

ネガティブ語辞書のデータ構造を示す図である。

【図14】

ネガティブ語トップワード辞書のデータ構造を示す図である。

【図15】

モダリティ表現辞書のデータ構造を示す図である。

【図16】

モダリティ表現トップワード辞書のデータ構造を示す図である。

【図17】

ネガティブ語候補抽出手順を示すフローチャートである。

【図18】

ネガティブ語辞書作成手順を示すフローチャートである。

【図19】

モダリティ表現候補抽出手順を示すフローチャートである。

【図20】

モダリティ表現辞書作成手順を示すフローチャートである。

【図21】

ネガティブ表現およびモダリティ表現の抽出手順を示すフローチャートである

【符号の説明】

101 : CPU

102 : 入力装置

103 : 表示装置

104 : コールセンタ応答履歴データベース

105 : シソーラスブラウジング用データ格納部

106 : 文書保存フォルダ

107 : 低頻度知識抽出用データ格納部

108：メモリ

1051：関連シソーラス格納部

1052：タームベクトル格納部

1053：およびシソーラス概観格納部

1071：ネガティブ文字辞書

1072：ネガティブ語辞書

1073：ネガティブ語ストップワード辞書

1074：モダリティ表現辞書

1075：モダリティ表現ストップワード辞書

1081：シソーラスブラウジング用データ生成処理手段

1082：シソーラスブラウジング処理手段

1083：文書検索手段

1084：ネガティブ語候補抽出手段

1085：ネガティブ語辞書作成手段

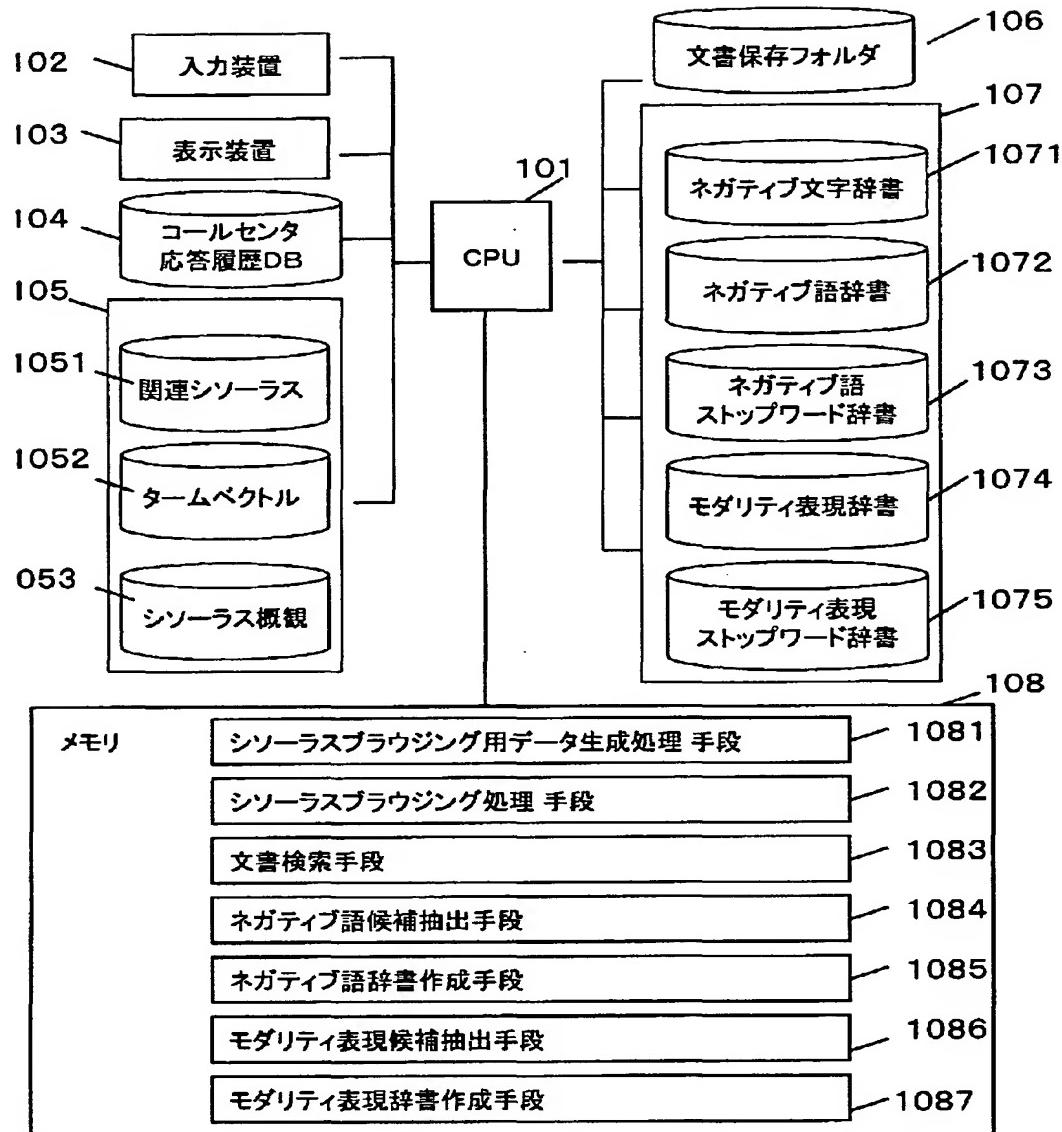
1086：モダリティ表現候補抽出手段

1087：モダリティ表現辞書作成手段。

【書類名】 図面

【図1】

図1



【図2】

図2

コールセンター応答履歴データベースのデータ構造

問合せID	応答履歴メモ	検索フラグ	分類フラグ
1			
2			
3			
4			
5			
6			

【図3】

図3

関連シソーラス格納部のデータ構造

#	タームX	タームY	関速度
1			
2			
3			
4			
5			
6			

【図4】

図4

タームベクトル格納部のデータ構造

#	問合せID	重要タームリスト
1	1	
2	2	
3	3	
4	4	
5	5	
6	6	

【図5】

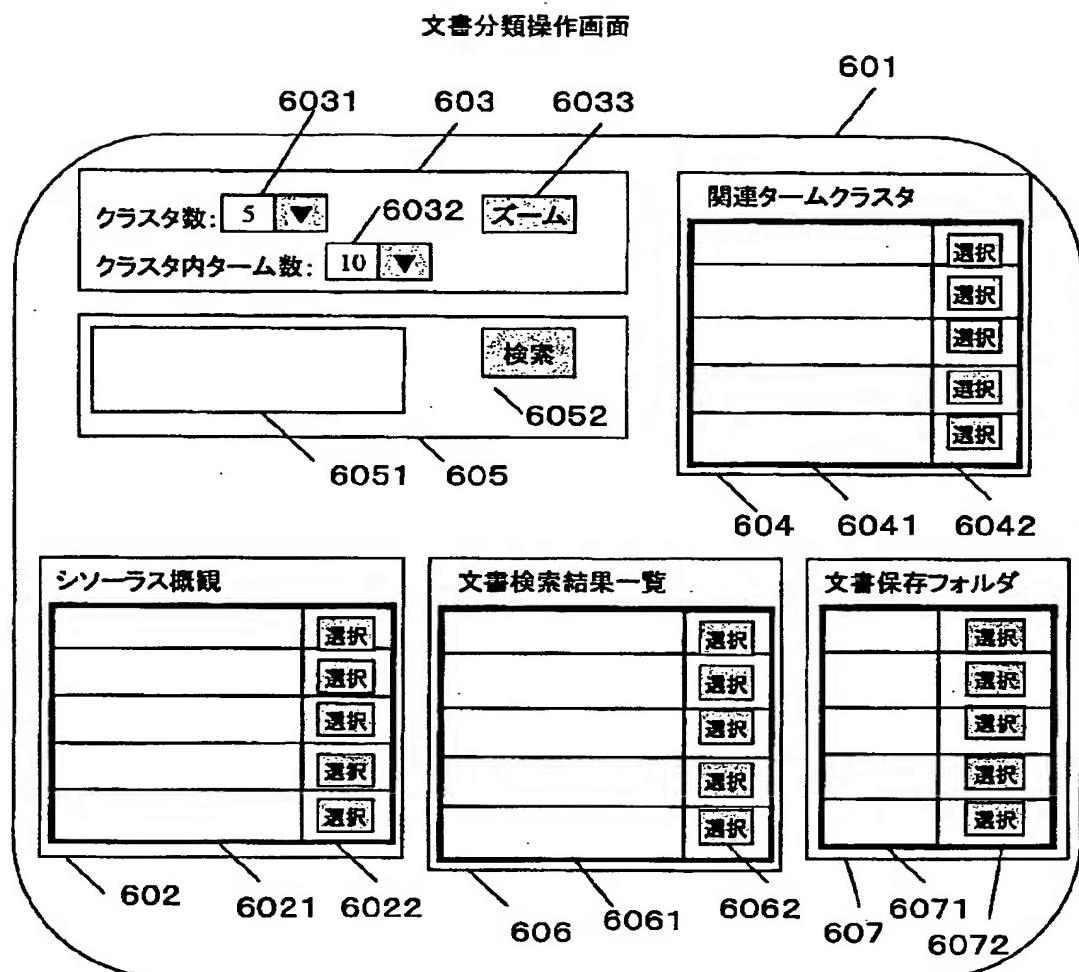
図5

シソーラス概観格納部のデータ構造

タームグループ番号	タームリスト
1	
2	
3	
4	
5	
6	

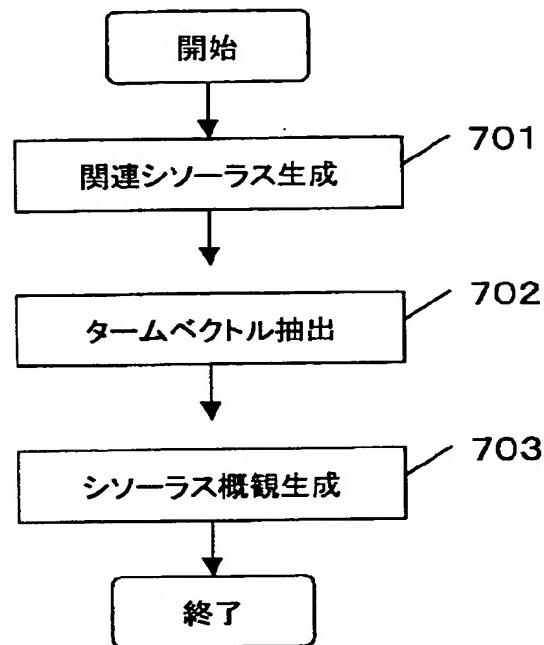
【図6】

図6



【図7】

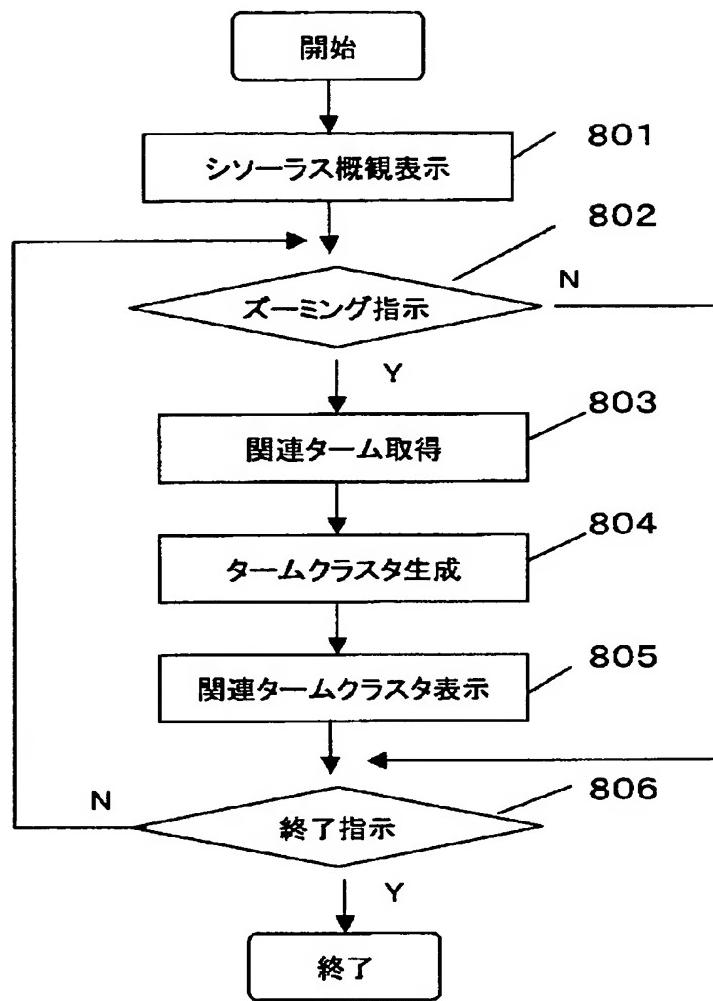
図7
シソーラスブラウジング用データ生成処理手順



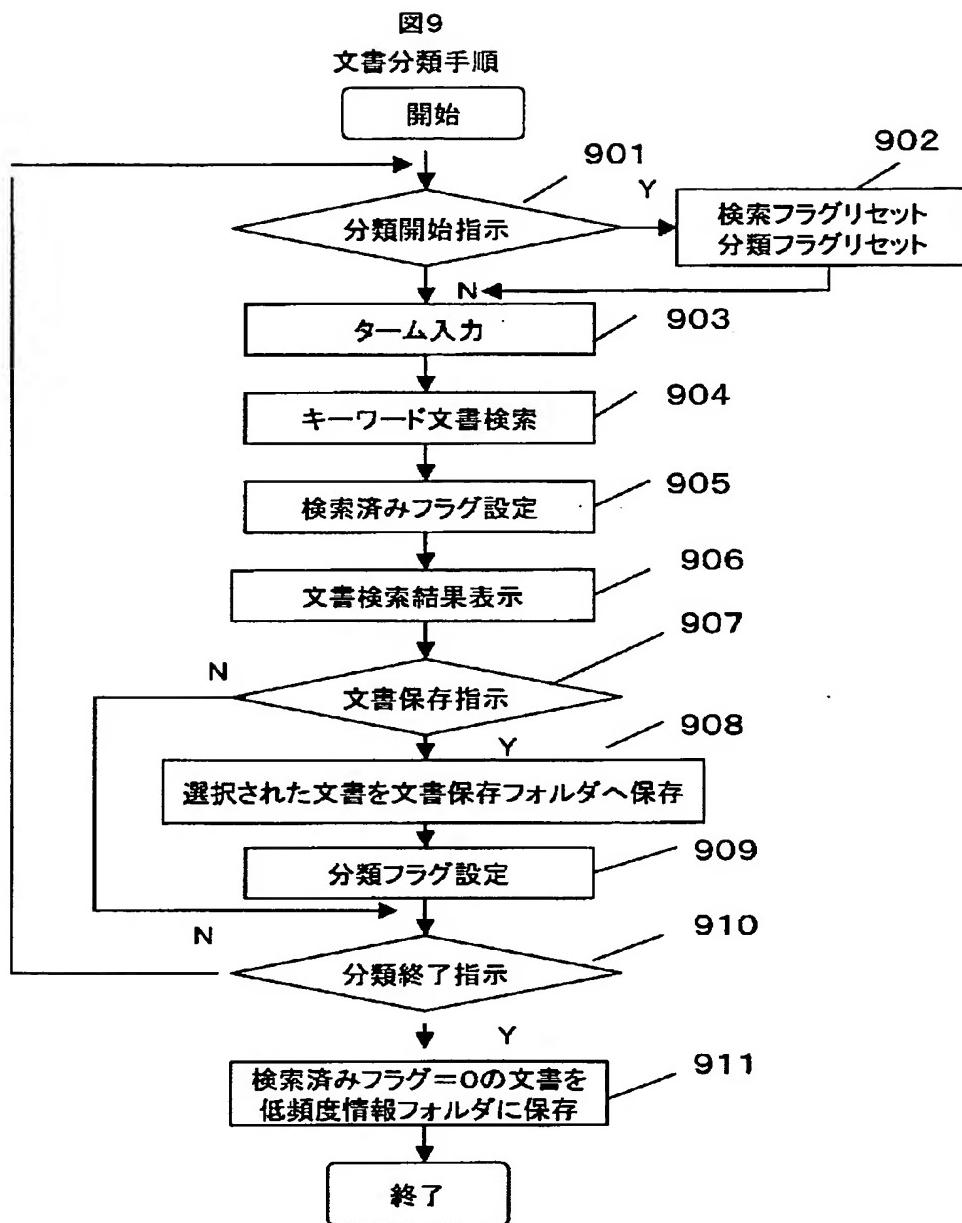
【図8】

図8

シソーラスブラウジング処理手順



【図9】



【図10】

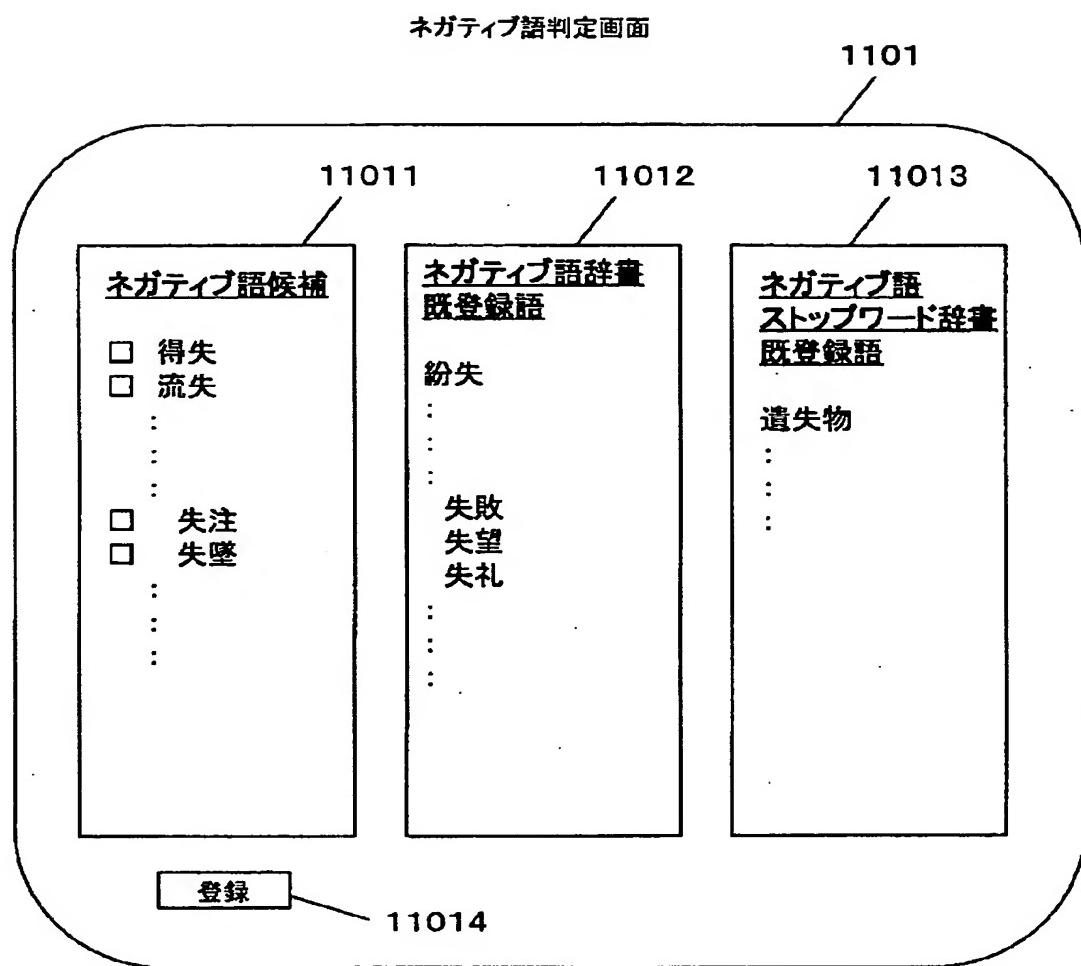
図10

文書保存フォルダのデータ構造

1141 フォルダID	1142 フォルダ名	1143 問合せIDリスト
1		
2		
3		
4		
5		
6		

【図11】

図11



【図12】

図12

ネガティブ文字辞書のデータ構造

ID	ネガティブ 文字	ネガティブ度	ネガティブ語辞 書登録語数	ネガティブ語 ストップワード辞書 登録語数
:				
35	失	0.8	32	4
:				
42	遅	0.7	5	0
:				
56	負	0.6	14	4

【図13】

図13

ネガティブ語辞書のデータ構造

ID	ネガティブ語	ネガティブ度
:		
131	紛失	0.8
132	失敗	0.8
133	失望	0.8
134	失礼	0.8
:		

【図14】

図14

ネガティブ語ストップワード辞書のデータ構造

ID	ネガティブ語 ストップワード
:	
12	遺失物
:	
45	勝負
:	
:	

【図15】

図15

モダリティ表現辞書のデータ構造

10741 10742 10743 10744

ID	モダリティ表現	品詞	モダリティ
25	くれないのか	助動詞	要望(非難)
:			
35	そもそも	副詞	批判
:			
42	なんなか	助動詞句	質問(非難)
:			
82	欲しい	助動詞	要望

【図16】

図16

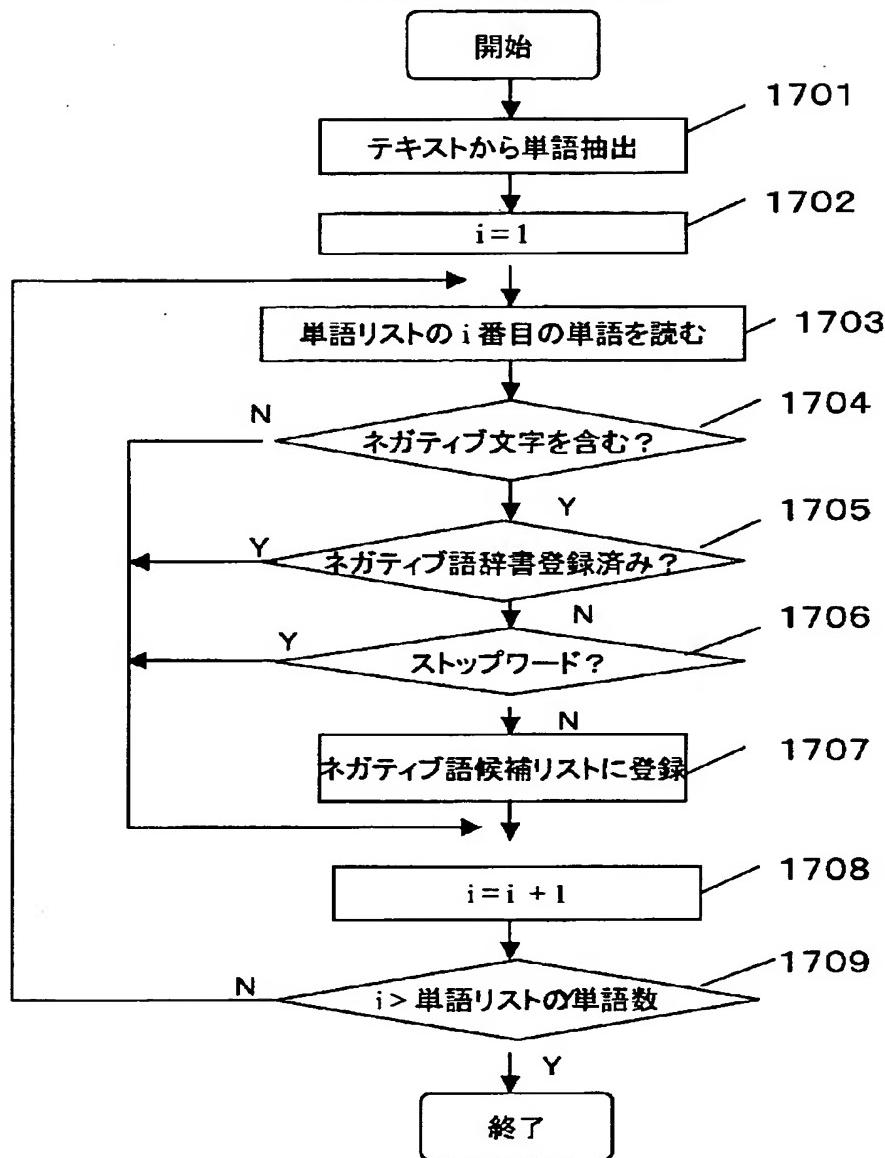
モダリティ表現ストップワード辞書のデータ構造

ID	モダリティ表現 ネガティブ語 ストップワード	品詞
:		
12	いつも	副詞
:		
45		
:		
:		

【図17】

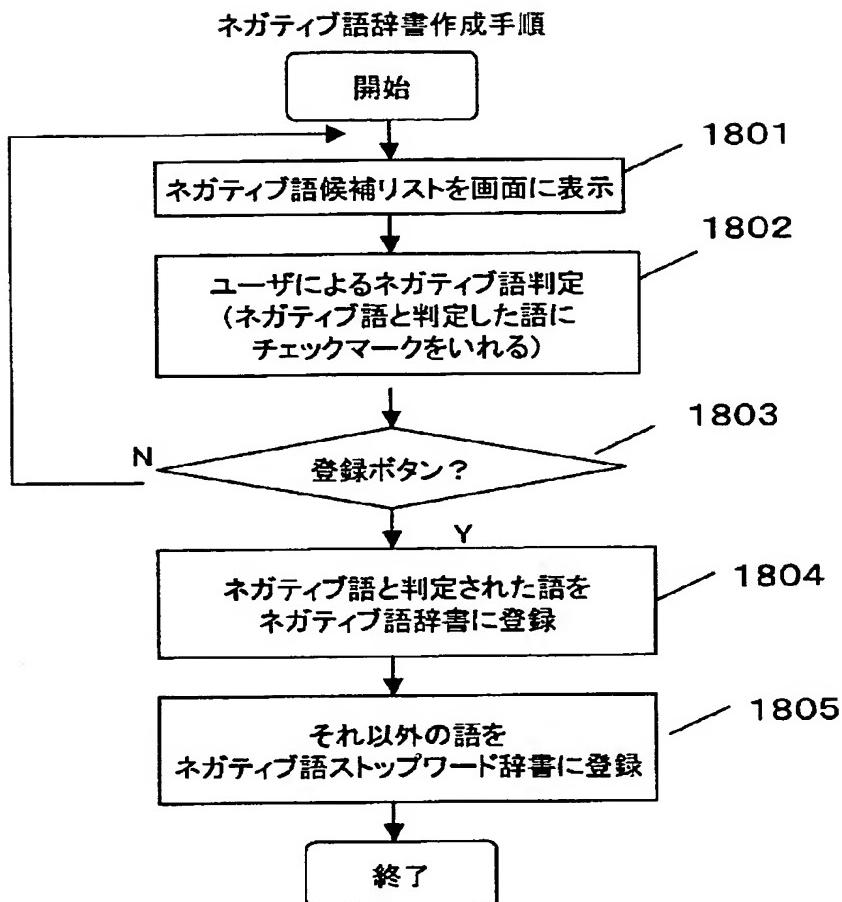
図17

ネガティブ語候補抽出手順



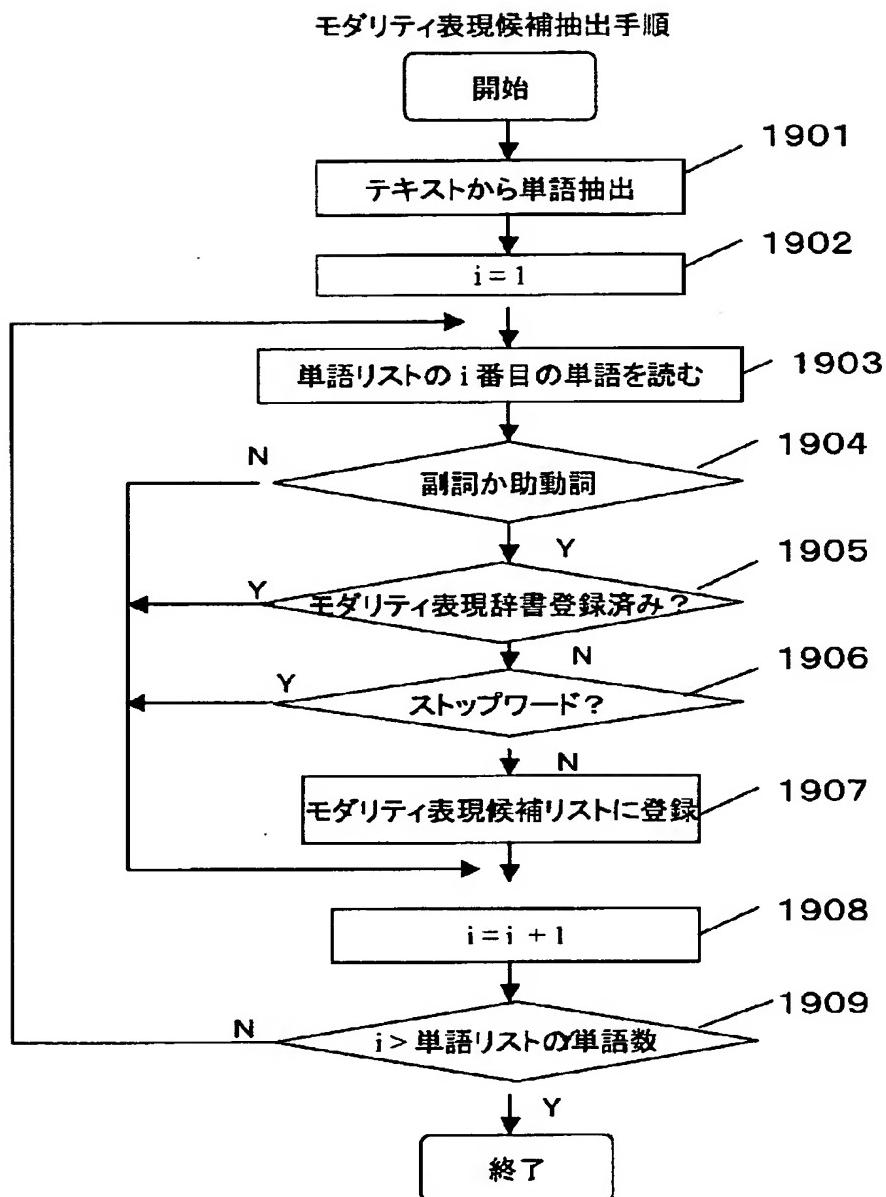
【図18】

図18



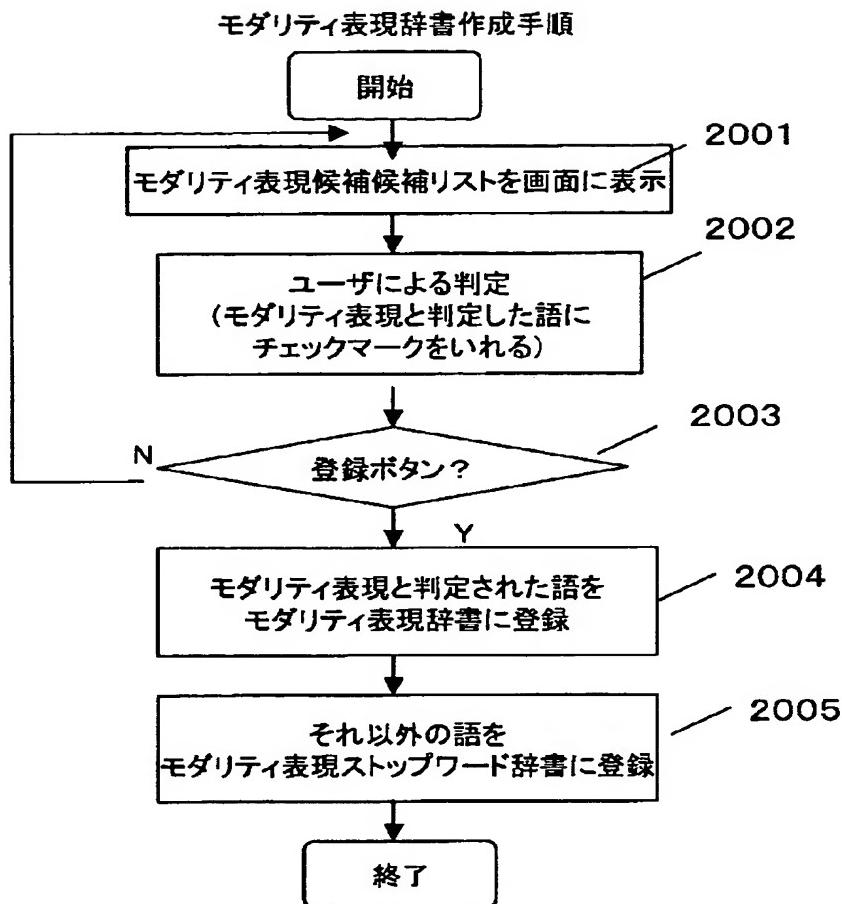
【図19】

図19

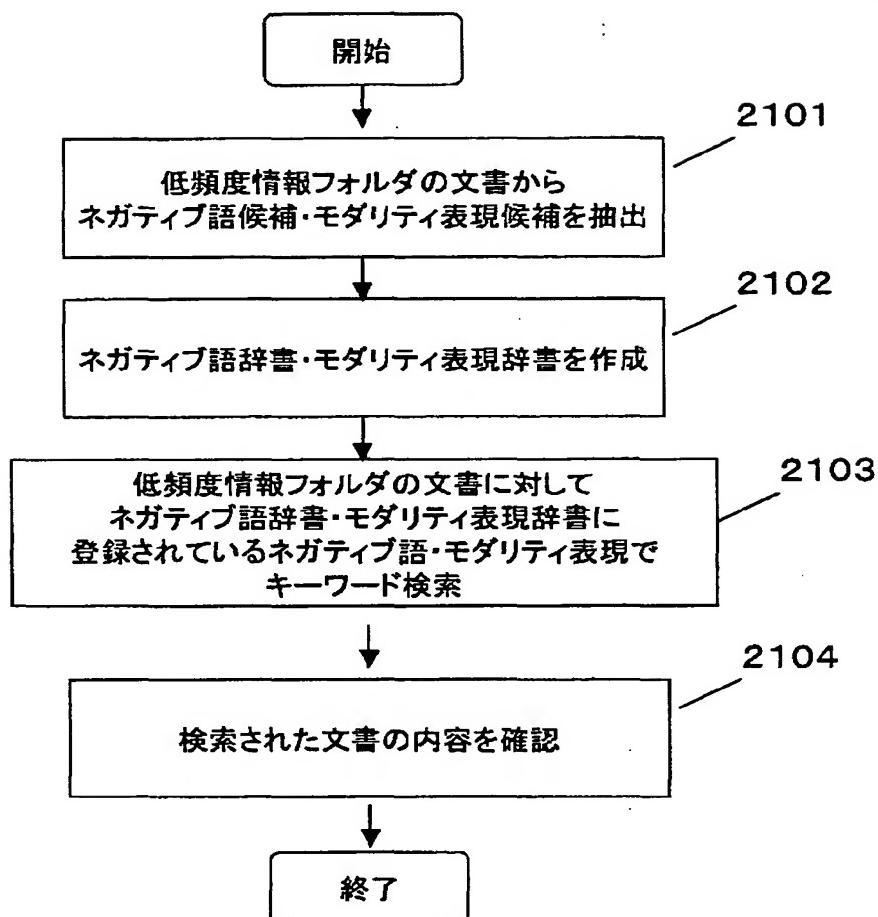


【図20】

図20



【図21】

図21
低頻度知識抽出手順

【書類名】 要約書

【要約】 テキスト分析支援方法において、高頻度情報と低頻度情報に分けるステップを設け、それぞれに適した分析方法を適用する。低頻度情報に対しては、ネガティブ表現、モダリティ表現を抽出することにより、リスク管理上有用な知識の抽出を支援する。

【課題】 従来のキーワードによる文書分類技術は、高頻度知識の抽出・分類に適しているが、コールセンターの応答履歴からリスク管理上有用な情報や顧客の生の声を抽出するには、大量のありふれた情報を取り除いた中から真に有用な知識を抽出する必要がある。

【解決手段】 キーワードで検索した文書をフォルダに保存する機能を設け、キーワード検索により高頻度情報をフォルダに保存した後、残りの文書を低頻度情報のフォルダに保存する機能を設ける。低頻度情報からリスク管理上有用な知識を抽出する手段として、ネガティブな表現や心的態度を表すモダリティ表現を抽出する機能を設ける。

【選択図】 図1

認定・付加情報

特許出願の番号	特願2002-341671
受付番号	50201780063
書類名	特許願
担当官	第七担当上席 0096
作成日	平成14年11月27日

<認定情報・付加情報>

【提出日】 平成14年11月26日

次頁無

出願人履歴情報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住所 東京都千代田区神田駿河台4丁目6番地
氏名 株式会社日立製作所